ABSTRACT OF THE DISCLOSURE

Techniques are provided for predicting the usage of a document collection given proximal cue information in the documents, a starting point and the user's information needs. A document collection topology matrix is created indicating links between document content portions. The link entry documents are analyzed for proximal cue words based on link URL, surrounding text and title. For image links, the connected to document information may also be used. Proximal cue words are added to a matrix relating proximal cue words and links. The proximal scent matrix indicates a similarity between the user's information need and the proximal cue word matrix. A distal scent information matrix is also calculated using distal document information and combined with the proximal scent matrix. Spreading activation is then applied to the resulting matrix using the starting location for a requested number of iterations and resulting in a predicted usage of the document collection.